Mushroom Classification

(Machine Learning Project Report)

Project Member:
Amardeep Kumar

Project Details:

- Problem Statement

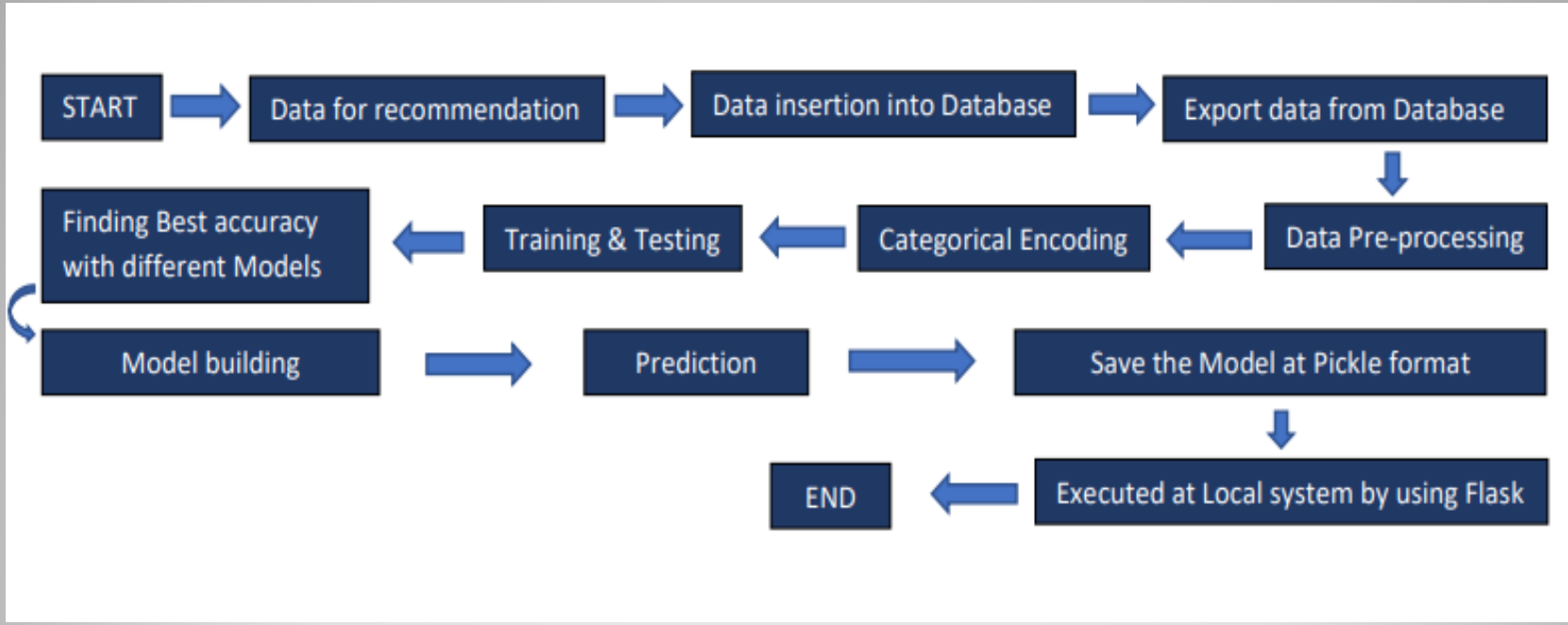
The Audubon Society Field Guide to North American Mushrooms contains descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family Mushroom (1981). Each species is labelled as either definitely edible, definitely poisonous, or maybe edible but not recommended. This last category was merged with the toxic category. The Guide asserts unequivocally that there is no simple rule for judging a mushroom's edibility, such as "leaflets three, leave it be" for Poisonous Oak and Ivy.

The main goal is to predict which mushroom is poisonous & which is edible

Objective:

Development of a system to find out whether a given mushroom is poisonous or not poisonous. The main goal is to find whether a mushroom is edible or not after studying its physical noticeable properties.

Architecture:



Dataset Information:

The dataset includes categorical characteristics on 8,124 mushroom samples from various species of gilled mushrooms.

- The target variable assessed was a class distinction of 'edible' or 'poisonous'.
- The explanatory variables covered a range of descriptive and visual characteristics on the structure of each observed mushroom - such as, cap color, odor, ring number and stalk shape.

Model Training:

1. Training and Testing Dataset:

As here 80 % of dataset has been trained and 20% of dataset has been tested.

2. Finding Accuracy with different models:

All the supervised machine learning algorithm were used to classify the output such as Logistic regression, K-neighbour, SVC, Decision tree, Random forest, Gradient boosting classifier, etc. found accuracy with every models.

3. Model Building:

After checking accuracy with different model, model building was created with the best accuracy and saved the model in pickle format.

Conclusion:

Our tuned classification models all performed really well with the dataset. Logistic Regression, which had a score of 99% would normally be a great choice but given that the model predicted false negatives which could be deadly, and that the other tested models performed perfectly, the other models are much better suited to classify mushrooms. Since our models performed so well, it was clear to us that they were able to identify specific traits that greatly influenced the classification of an edible versus poisonous mushroom. And that was exactly what we were hoping for!